Supplementary Material for
"Generalized Linear Mixed Models with Gaussian Mixture Random Effects: Inference and Application"

Lanfeng Pan[a], Yehua Li[b,*], Kevin He[c], Yanming Li[c], Yi Li[c]

[a]*Amazon.com, Inc., Seattle, WA 98109, U.S.A.*
[b]*Department of Statistics, University of California at Riverside, 900 University Ave. Riverside, CA 92521, U.S.A.*
[c]*School of Public Health & Kidney Epidemiology and Cost Center, University of Michigan, Ann Arbor, MI 48109, U.S.A.*

In this supplementary paper, we provide details of the model fitting algorithm in Section S1 and a simulation procedure in Section S2 to evaluate the asymptotic distribution in Proposition 5. In Section S3, we also provide some additional simulation results.

S1. Model fitting using EM algorithm

We now provide the detailed algorithm to maximize the penalized likelihood in Section 2.

S1.1. E-Step with Gauss-Hermite quadrature approximation

At the tth iteration of the algorithm, given the parameter value $\boldsymbol{\theta}^{(t-1)}$ from the previous iteration, we first evaluate the following loss function at the E-step

$$Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t-1)}) = \sum_{i=1}^{n} E\big[\ell_{i,comp}(\theta; \boldsymbol{Y}_i, \boldsymbol{X}_i, \gamma_i, \boldsymbol{L}_i) \mid \boldsymbol{Y}_i, \boldsymbol{X}_i, \boldsymbol{\theta}^{(t-1)}\big] + \sum_{c=1}^{C} p_n(\sigma_c^2; \widehat{\sigma}_{pilot}^2) \tag{S.1}$$

where

$$\begin{aligned} & E\left[\ell_{i,comp}(\theta; \boldsymbol{Y}_i, \boldsymbol{X}_i, \gamma_i, \boldsymbol{L}_i) \mid \boldsymbol{Y}_i, \boldsymbol{X}_i, \boldsymbol{\theta}^{(t-1)}\right] \\ = \quad & \sum_{c=1}^{C} \int \log f(\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \gamma; \boldsymbol{\theta}_y) f(\gamma, L_{ic} = 1 \mid \boldsymbol{X}_i, \boldsymbol{Y}_i; \boldsymbol{\theta}^{(t-1)}) d\gamma \\ & + \sum_{c=1}^{C} \int \log f_c(\gamma \mid \mu_c, \sigma_c) f(\gamma, L_{ic} = 1 \mid \boldsymbol{X}_i, \boldsymbol{Y}_i; \boldsymbol{\theta}^{(t-1)}) d\gamma \\ & + \sum_{c=1}^{C} \log \pi_c \int f(\gamma, L_{ic} = 1 \mid \boldsymbol{X}_i, \boldsymbol{Y}_i; \boldsymbol{\theta}^{(t-1)}) d\gamma, \end{aligned}$$

*Corresponding author
Email addresses: planfeng@amazon.com (Lanfeng Pan), yehuali@ucr.edu (Yehua Li), kevinhe@umich.edu (Kevin He), liyanmin@umich.edu (Yanming Li), yili@med.umich.edu (Yi Li)

Preprint submitted to Journal of Multivariate Analysis *September 3, 2019*

and $f(\gamma, L_{ic}=1 \mid \boldsymbol{X}_i, \boldsymbol{Y}_i; \boldsymbol{\theta}^{(t-1)})$ is

$$\frac{\pi_c^{(t-1)} f(\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \gamma; \boldsymbol{\theta}_y^{(t-1)}) \frac{1}{\sigma_c^{(t-1)}} \phi\left(\frac{\gamma-\mu_c^{(t-1)}}{\sigma_c^{(t-1)}}\right)}{\sum_{c=1}^C \pi_c^{(t-1)} \int f(\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \gamma; \boldsymbol{\theta}_y^{(t-1)}) \frac{1}{\sigma_c^{(t-1)}} \phi\left(\frac{\gamma-\mu_c^{(t-1)}}{\sigma_c^{(t-1)}}\right) d\gamma}.$$

Expectation for a function of a Gaussian random variable can be closely approximated by Gauss-Hermite quadrature:

$$\int h(\gamma) \frac{1}{\sigma} \phi\{(\gamma-\mu)/\sigma\} d\gamma \approx \frac{1}{\sqrt{\pi}} \sum_{m=1}^M w_m h(\gamma_m)$$

where $h(\gamma)$ is an integrable real valued function, $\gamma_m = \mu + \sqrt{2}\sigma d_m$, $d_1, \ldots, d_M$ are the Gauss-Hermite abscissas and $w_1, \ldots, w_M$ are the corresponding quadrature weights. We find in our numerical studies that using $M = 100$ quadrature points usually provides a close enough approximation. Denote $\gamma^{(c,m)} = \mu_c^{(t-1)} + \sqrt{2}\sigma_c^{(t-1)} d_m$. The Gauss-Hermite approximation for $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t-1})$ is

$$\begin{aligned}
\widehat{Q}(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t-1)}) &= \sum_{i=1}^n \frac{\sum_{c=1}^C \sum_{m=1}^M w_m \pi_c^{(t-1)} \log f(\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \gamma^{(c,m)}; \boldsymbol{\theta}_y) f(\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \gamma^{(c,m)}; \boldsymbol{\theta}_y^{(t-1)})}{\sum_{c=1}^C \sum_{m=1}^M w_m \pi_c^{(t-1)} f(\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \gamma^{(c,m)}; \boldsymbol{\theta}_y^{(t-1)})} \\
&\quad + \sum_{i=1}^n \frac{\sum_{c=1}^C \sum_{m=1}^M w_m \pi_c^{(t-1)} \left[-\frac{1}{2}\log 2\pi\sigma_c^2 - \frac{1}{2}\frac{(\gamma^{(c,m)}-\mu_c)^2}{\sigma_c^2}\right] f(\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \gamma^{(c,m)}; \boldsymbol{\theta}_y^{(t-1)})}{\sum_{c=1}^C \sum_{m=1}^M w_m \pi_c^{(t-1)} f(\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \gamma^{(c,m)}; \boldsymbol{\theta}_y^{(t-1)})} \\
&\quad + \sum_{i=1}^n \frac{\sum_{c=1}^C \sum_{m=1}^M w_m \pi_c^{(t-1)} \log \pi_c f(\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \gamma^{(c,m)}; \boldsymbol{\theta}_y^{(t-1)})}{\sum_{c=1}^C \sum_{m=1}^M w_m \pi_c^{(t-1)} f(\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \gamma^{(c,m)}; \boldsymbol{\theta}_y^{(t-1)})} + \sum_{c=1}^C p_n(\sigma_c^2; \widehat{\sigma}_{pilot}^2) \\
&= \sum_{i=1}^n \sum_{c=1}^C \sum_{m=1}^M \omega_{icm} \left\{\log f(\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \gamma^{(c,m)}; \boldsymbol{\theta}_y) - \frac{1}{2}\log 2\pi\sigma_c^2 - \frac{1}{2}\frac{(\gamma^{(c,m)}-\mu_c)^2}{\sigma_c^2} + \log\pi_c\right\} \\
&\quad - a_n \sum_{c=1}^C \{\widehat{\sigma}_{pilot}^2/\sigma_c^2 + \log(\sigma_c^2/\widehat{\sigma}_{pilot}^2) - 1\},
\end{aligned}$$

where

$$\omega_{icm} = \frac{w_m \pi_c^{(t-1)} f(\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \gamma^{(c,m)}; \boldsymbol{\theta}_y^{(t-1)})}{\sum_{c'=1}^C \sum_{m'=1}^M w_{m'} \pi_{c'}^{(t-1)} f(\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \gamma^{(c',m')}; \boldsymbol{\theta}_y^{(t-1)})}. \tag{S.2}$$

S1.2. M-Step

In the M-step, we maximize $\widehat{Q}(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t-1)})$ with respect to $\boldsymbol{\theta}$, and update different components of $\boldsymbol{\theta}$ by

$$\begin{aligned}
\pi_c^{(t)} &= \frac{1}{n}\sum_{i=1}^n \sum_{m=1}^M \omega_{icm}, \qquad \mu_c^{(t)} = \frac{\sum_{i=1}^n \sum_{m=1}^M \gamma^{(c,m)} \omega_{icm}}{\sum_{i=1}^n \sum_{m=1}^M \omega_{icm}}, \\
(\sigma_c^2)^{(t)} &= \frac{\sum_{i=1}^n \sum_{m=1}^M (\gamma^{(c,m)} - \mu_c^{(t)})^2 \omega_{icm} + 2a_n \widehat{\sigma}_{pilot}^2}{\sum_{i=1}^n \sum_{m=1}^M \omega_{icm} + 2a_n},
\end{aligned}$$

and obtain $\boldsymbol{\theta}_y^{(t)}$ by maximizing $\sum_{i=1}^n \sum_{c=1}^C \sum_{m=1}^M \omega_{icm} \log f(\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \gamma^{(c,m)}; \boldsymbol{\theta}_y)$ using iteratively reweighted least squares.

S1.3. Stopping rule and random effect prediction

Following Booth and Hobert [1], we stop the EM algorithm at iteration t if

$$\max_l \frac{|\theta_l^{(t)} - \theta_l^{(t-1)}|}{|\theta_l^{(t-1)}| + 0.001} < 0.005,$$

where θ_l is the lth entry in $\boldsymbol{\theta}$.

At convergence, the weight ω_{icm} can be used to calculate some other quantities of interest, such as the marginal likelihood, the posterior probability of γ_i belonging to the cth component and posterior mean of γ_i . For example, we predict γ_i by its posterior mean

$$\int \gamma f(\gamma \mid \boldsymbol{Y}_i, \boldsymbol{X}_i, \boldsymbol{\theta}) d\gamma = \frac{\sum_{c=1}^{C} \pi_c \int \gamma f(\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \gamma; \boldsymbol{\theta}_y) \phi\{(\gamma - \mu_c)/\sigma_c\}/\sigma_c d\gamma}{\sum_{c=1}^{C} \pi_c \int f(\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \gamma; \boldsymbol{\theta}_y) \phi\{(\gamma - \mu_c)/\sigma_c\}/\sigma_c d\gamma}.$$

Using the Gauss-Hermite approximation, the posterior mean is approximated by

$$\widehat{\gamma}_i = \sum_{c=1}^{C} \sum_{m=1}^{M} \gamma^{(c,m)} \omega_{icm}, \tag{S.3}$$

where ω_{icm} is defined in (S.2) evaluated at $\widehat{\boldsymbol{\theta}}$.

To obtain reasonable initial values for $\boldsymbol{\theta}_y$ and $\boldsymbol{\theta}_\gamma$, we first run a generalized linear mixed model assuming γ_i's are i.i.d. normal. We use the estimated fixed effects as initial values for $\boldsymbol{\theta}_y$, fit a Gaussian mixture model on the predicted values $\widehat{\boldsymbol{\gamma}}$ and use the results as the initial values for $\boldsymbol{\theta}_\gamma$.

S2. Simulation Approach for the Asymptotic Distribution in Proposition 5

We use the following procedure to simulate the asymptotic distribution in Proposition 5 under the hypothesis $H_0 : C_0 = C$.

Step 0. Fit a C-component latent Gaussian mixture model and obtain the reduced model estimator $\widehat{\boldsymbol{\theta}}_{red}$.

Step 1. Calculate $\tilde{\boldsymbol{s}}_i = (\boldsymbol{s}_{\boldsymbol{\eta},i}^\top, \tilde{\boldsymbol{s}}_{\boldsymbol{\lambda},i}^\top)^\top$ with $\tilde{\boldsymbol{s}}_{\lambda,i} = \{(\boldsymbol{s}_{\lambda,i}^{(1)})^\top, \ldots (\boldsymbol{s}_{\lambda,i}^{(C)})^\top\}^\top$, where $\boldsymbol{s}_{\eta,i}$ and $\boldsymbol{s}_{\lambda,i}^{(c)}$, $c = 1, \ldots, C$, are the score functions for the restricted full models defined in (11) evaluated at $\widehat{\boldsymbol{\theta}}_{red}$. Let

$$\tilde{\boldsymbol{I}} = \frac{1}{n} \sum_{i=1}^{n} \tilde{\boldsymbol{s}}_i (\tilde{\boldsymbol{s}}_i)^T = \begin{pmatrix} \boldsymbol{I}_\eta & \tilde{\boldsymbol{I}}_{\eta\lambda} \\ \tilde{\boldsymbol{I}}_{\lambda\eta} & \tilde{\boldsymbol{I}}_\lambda \end{pmatrix}$$

be the sample version of $\tilde{\boldsymbol{\mathcal{I}}} = E\tilde{\boldsymbol{s}}_i\tilde{\boldsymbol{s}}_i^T$, and calculate $\tilde{\boldsymbol{I}}_{\lambda|\eta} = \tilde{\boldsymbol{I}}_\lambda - \tilde{\boldsymbol{I}}_{\lambda\eta}\boldsymbol{I}_\eta^{-1}(\tilde{\boldsymbol{I}}_{\lambda\eta})^T$. To improve numerical stability, we check if $\tilde{\boldsymbol{I}}$ is an ill conditioned matrix. If so, set the eigenvalues with small absolute values to be a small positive number.

Step 2. Generate random a vector

$$\boldsymbol{s} = \left\{(\boldsymbol{s}^{(1)})^\top, \ldots, (\boldsymbol{s}^{(C)})^\top\right\}^\top \sim N(0, \tilde{\boldsymbol{I}}_{\lambda|\eta}).$$

Let $\boldsymbol{I}_{\lambda|\eta}^{(c)}$ be the sub diagonal matrix of $\tilde{\boldsymbol{I}}_{\lambda|\eta}$ corresponding to $\boldsymbol{s}^{(c)}$. Then

$$T_C^* = \max\left\{(\boldsymbol{s}^{(c)})^T (\boldsymbol{I}_{\lambda|\eta}^{(c)})^{-1} \boldsymbol{s}^{(c)}, c = 1, \ldots, C\right\}$$

has the same asymptotic distribution as $T_C(\tau)$ and $\widetilde{T}_C$.

Step 3. Repeat Step 2 a large number of times and use the empirical distribution of T_C^* to approximate the asymptotic distribution of $\widetilde{T}_C$.

S3. Additional Simulation Results

Recall that Models 1 and 2 in the simulation study are latent Gaussian mixture models with 2 and 3 mixture components respectively. Tables S.1 and S.2 are the estimation results for the model parameters, when the model is misspecified as the classic GLMM with a homogeneous Gaussian random effect.

Table S.1: Additional simulation results for Model 1, when the model is misspecified as GLMM with a Gaussian random effect. Results are based on 200 replications.

	Mean	Std
π_1	1.0000	0.0000
μ_1	-1.2716	0.1208
σ_1	2.3397	0.0722
β_1	1.0003	0.0203
β_2	1.0010	0.0204

Table S.2: Additional simulation results for Model 2, when the model is misspecified as GLMM with a Gaussian random effect. Results are based on 200 replications.

	Mean	Std
π_1	1.0000	0.0000
μ_1	-0.8085	0.1745
σ_1	3.4584	0.1371
β_1	1.0057	0.0240
β_2	1.0082	0.0218

[1] J. G. Booth, J. P. Hobert, Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm, JJ. R. Stat. Soc. Ser. B. Stat. Methodol. 61 (1999) 265–285.